A General Variational VM-Update for Unconstrained Optimization

Abbas Y. AL-Bayati[1] , Basim A. Hassan[1] , Zeyad M. Abdullah[2]
[1] *Department of Mathematics , College of Computers Sciences and Mathematics , University of Mosul, Mosul , Iraq*
[2] *Department of Computers , College of Computers Sciences and Mathematics , University of Tikrit, Tikrit , Iraq*

(Received: 11 / 5 / 2010---- Accepted: 26 / 10 / 2011)

Abstract
In this paper, we have proposed a new variational Variable Metric (VM) method for solving unconstrained optimization problems . Experimental results indicate that the new proposed method was efficient than the standard BFGS method .

Introduction
Consider the Variable Metric (VM) method whose iteration step has the form :

$$x_{k+1} = x_k + \alpha_k d_k \quad ....(1)$$

Where

$$d_k = -H_k g_k \quad ....(2)$$

Here x_k and x_{k+1} are old and new vectors of variables, respectively. Direction vector d_k and positive step size α_k are chosen so that

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta_1 \alpha_k d_k^T g_k \quad ....(3)$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \delta_2 d_k^T g_k \quad ....(4)$$

with $0 < \delta_1 \prec 1/2$ and $\delta_1 \prec \delta_2 \prec 1$, where $f(x_k + \alpha_k d_k)$ and $f(x_k)$ are old and new values of the objective function and $g(x_k + \alpha_k d_k)$ and g_k are old and new gradient of the objective function, respectively. H_k is a symmetric positive definite approximation of the inverse Hessian matrix that is constructed iteratively using the formula

$$H_{k+1}^{BFGS} = H_k - \frac{H_k y_k v_k^T + v_k y_k^T H_k}{v_k^T y_k} + \frac{v_k v_k^T}{v_k^T y_k}\left[1 + \frac{y_k^T H_k y_k}{v_k^T y_k}\right] \quad ....(5)$$

This is the so called BFGS method **[7]**.

The general strategy of self-scaling VM-method is to scale the Hessian approximation matrix H_k before it is updated at each iteration . This is to avoid large difference in the eigenvalues of the approximated Hessian of the objective function. Self-scaling VM methods were introduced by Oren, see **[5,6]**. The Hessian approximation matrix H_k can be updated according to a self-scaling BFGS update of the form:

$$H_k^{Oren} = \xi_k\left(H_k - \frac{H_k y_k v_k^T + v_k y_k^T H_k}{v_k^T y_k}\right) + \frac{v_k v_k^T}{v_k^T y_k}\left[1 + \frac{y_k^T H_k y_k}{v_k^T y_k}\right] \quad ....(6)$$

where

$$\xi_k = \frac{v_k^T y_k}{y_k^T H_k y_k} \quad ....(7)$$

The matrix H_{k+1} satisfies the Quasi-Newton (QN) condition

$$H_{k+1} y_k = v_k \quad ....(8)$$

where $y_k = g_{k+1} - g_k$, $v_k = x_{k+1} - x_k$, ξ_k is a scalar see **[1]** .

Variational Formulation
In **[4]**, Greenstadt derives VM methods, using a classical variational approach. Specifically, iterative formulas are developed for updating the matrix H_k , the inverse of the VM, where H_k is an approximation to the inverse Hessian $G^{-1}(x_k)$ of the function being minimized . Using the iterative formula :

$$H_{k+1} = \xi_k (H_k + \theta_k w_k w_k^T) + E_k \quad ....(9a)$$

where

$$w_k = (y_k^T H_k y_k)^T \left[\frac{v_k}{v_k^T y_k} - \frac{H_k y_k}{y_k^T H_k y}\right] \quad ....(9b)$$

Note that $w^T H w = a$ for a given value of $-1/a < \theta \le 1$. For more details see **[2]**.

To provide revised estimates to the inverse Hessian at each step , solving the correction term E_k that minimizes the norm yields

$$N(E_k) = Tr(W E_k W E_k^T) \quad ....(10)$$

subject to the conditions

$$E_k^T - E_k = 0 \quad ....(11)$$

And

$$v_k = H_{k+1} y_k = (\xi_k (H_k + \theta_k w_k w_k^T) + E_k) y_k \quad ....(12)$$

which reduces to :

$$E_k y_k = v_k - \xi_k (H_k + \theta_k w_k w_k^T) y_k \equiv r_k \quad ....(13)$$

where W is a positive definite symmetric matrix . In **[8]** Consider W -norm given by

$$\|E_k\|_W^2 = Tr(WEWE^T) \quad ....(14)$$

To ensure affine invariance, $Wv = y$ is required in defining the norm **[3]** .

In the remainder of this derivation, we shall ignore the subscript k .

We shall solve this constrained minimization problem by the use of Lagrange multipliers. We form the composite function Φ as follows :

$$\Phi \equiv \frac{1}{2} Tr(WEWE^T) + \lambda^T (Ey - r) + Tr\left[\Gamma(E - E^T)\right] . \quad ....(15)$$

We also note that

$$\lambda^T (Ey - r) = Tr\left[(Ey - r)\lambda^T\right] . \quad ....(16)$$

Hence, we have

$$\frac{\partial \Phi}{\partial E} = WEW + \lambda y^T + \Gamma^T - \Gamma = 0 \quad ....(17)$$

so that

$$E=-M\left[\lambda y^T+\Gamma^T-\Gamma\right]M \quad ....(18)$$

where $M=W^{-1}$. Since M is symmetric .

Subtracting E^T from E should give zero, so that

$$E-E^T=-M\left[\lambda y^T-y\lambda^T+2\Gamma^T-2\Gamma\right]M=0 \quad ....(19)$$

and we have

$$\Gamma^T-\Gamma=\frac{1}{2}\left[y\lambda^T-\lambda y^T\right] \quad ....(20)$$

Subtracting this into (18) gives

$$\begin{aligned} E &=-M\left[\lambda y^T+\frac{1}{2}(y\lambda^T-\lambda y^T)\right]M \\ &=-\frac{1}{2}M\left[y\lambda^T+\lambda y^T\right]M \end{aligned} \quad ....(21)$$

Now we take note of the condition ; Eq. (13) :

$$Ey-r=-\frac{1}{2}M\left[y\lambda^T+\lambda y^T\right]My-r=0. \quad ....(22)$$

multiplying by $2W$, we have

$$\left[y\lambda^T+\lambda y^T\right]My+2Wr=0 \quad ....(23)$$

From which we solve for the λ which is free from the inner product. The result is

$$\lambda=-(y^T My)^{-1}\left[2Wr+y(\lambda^T My)\right] \quad ....(24)$$

we now multiply by $y^T M$ to obtain :

$$y^T M\lambda=-(y^T My)^{-1}\left[2Wr+(y^T My)((\lambda^T My)\right] \quad ....(25)$$

and, since $y^T M\lambda$ is the same as $\lambda^T My$, we can solve for $\lambda^T My$. The result is

$$\lambda^T My=-(y^T My)^{-1}(y^T r) \quad ....(26)$$

we now substitute this back into (14) to obtain :

$$\begin{aligned} \lambda &= -(y^T My)^{-1}\left[2Wr-(y^T My)^{-1}(y^T r)y\right] \\ &= (y^T My)^{-2}(y^T r)y-2(y^T My)^{-1}Wr \end{aligned} \quad ....(27)$$

and we are in a position to replace λ in Eq. (21) . We then have for E

$$E=\frac{1}{y^T My}\left\{ry^T M+Myr^T-\left(\frac{y^T r}{y^T My}\right)Myy^T M\right\} \quad ....(28)$$

and, finally, replacing r by $v-\xi(H_k+\theta ww^T)y$, we obtain

$$E=\frac{1}{y^T My}\left\{\begin{matrix} vy^T M+Myv^T-\xi(H_k+\theta ww^T)yy^T M-\xi Myy^T(H_k+\theta ww^T) \\ -\frac{1}{y^T My}\left(y^T v-\xi y^T(H_k+\theta ww^T)y\right)Myy^T M \end{matrix}\right\}$$

.........(29)

which is our final formula for E . Since $Wv=y \Rightarrow My=v$

$$E=\frac{1}{y^T v}\left\{\begin{matrix} vv^T+vv^T-\xi(H_k+\theta ww^T)yv^T-\xi vy^T(H_k+\theta ww^T) \\ -\frac{1}{y^T v}\left(y^T v-\xi y^T(H_k+\theta ww^T)y\right)vv^T \end{matrix}\right\}$$

.........(30)

we shall also refer to the new update which :

$$H_{k+1}=\xi_k(H_k+\theta ww^T)-\frac{\xi(H_k+\theta ww^T)yv^T+\xi vy^T(H_k+\theta ww^T)}{y^T v}+\frac{vv^T}{y^T v}\left\{1+\frac{1}{y^T v}\left(\xi y^T(H_k+\theta ww^T)y\right)\right\}$$

.........(31)

where

$$\xi_k=\frac{v_k^T y_k}{y_k^T(H_k+\theta ww^T)y_k}, \quad -1/a<\theta\le 1 \quad ....(32)$$

And

$$w_k=(y_k^T H_k y_k)^T\left[\frac{v_k}{v_k^T y_k}-\frac{H_k y_k}{y_k^T H_k y}\right] \quad ....(33)$$

A new Algorithm

Step 0: Choose an initial point $x_1\in R^n$, set $k=1$.

Step 1: If the stopping criterion is satisfied stop :

$\|g_{k+1}\|<\varepsilon$, $\varepsilon=1*10^{-4}$,

Step 2: Solve $d_k=-H_k g_k$ to obtain a search direction d_k .

Step 3: Find a step size α_k which satisfy the rules (3) and (4)

Step 4:Generate a new iteration point by $x_{k+1}=x_k+\alpha_k d_k$ and calculate .

the new updating formula (31) .

Step 5: Set $k=k+1$ and go to Step 1 .

Numerical Results

This section was devoted to numerical experiments. Our purpose was to check whether the new algorithm provide improvements on the corresponding standard BFGS-algorithm. The programs were written in Fortran 90. The test functions were commonly used for unconstrained test problems with standard starting points and a summary of the results of these test functions was given in Table (3.1). The same line search was employed in each algorithm, this was the cubic interpolation technique. The initial inverse approximation was $H_0=I$. The stopping criterion was taken to be $\|g_{k+1}\|\le\varepsilon$. We have used the dimension of the problem (n), n=100, 1000 .

We tabulate for comparison of these algorithms, the number of function evaluations (NOF) and the number of iterations (NOI) .

Table (3-1)

shows the computational results, where the columns have the following meanings :

Test functions : the name of the test problems .

NOI : Number of iterations .

NOF: Number of function evaluations .

Table (3.1)
Comparison between the New and the standard BFGS methods

Test functions	N	New NOF (NOI)	BFGS NOF (NOI)
Rosen	100	72 (20)	631 (234)
	500	72 (20)	1334 (545)
	1000	74 (21)	1445 (659)
Wood	100	96 (26)	630 (228)
	500	96 (26)	2026 (755)
	1000	96 (26)	2940 (1193)
Non-diagonal	100	79 (24)	213 (91)
	500	83 (24)	211 (92)
	1000	85 (25)	393 (140)
Wolfe	100	108 (36)	125 (62)
	500	116 (39)	141 (70)
	1000	125 (42)	165 (82)
Cubic	100	49 (12)	87 (37)
	500	49 (12)	111 (50)
	1000	49 (12)	100 (45)
Powell	100	89 (22)	96 (38)
	500	164 (43)	100 (40)
	1000	166 (42)	135 (53)
Cantrell	100	46 (13)	45 (11)
	500	46 (13)	56 (13)
	1000	57 (16)	56 (13)
Miele	100	81 (27)	84 (28)
	500	81 (27)	96 (32)
	1000	81 (27)	95 (32)
Total		2060 (595)	11315 (4543)

Conclasions and Discussions

In this paper, we have proposed a new vartional VM-type method for solving unconstrained minimization problems. The computational experiments show that the new approaches given in this paper are successful. We claim that the new method is better than the original formula. Namely, for the new method there are about 81.79 % improvement in NOI & there are about 86.90 % improvement in NOF overall, the calculations & for different dimensions .

Relative efficiency of the different methods discussed in the paper .

Tools	BFGS	new
NOI	100 %	18.20 %
NOF	100 %	13.90 %

However, for the 24 different cases the new algorithm has beresults in 20/24 % while BFGS has 4/24 % of the cases .

Appendix

1.*Cubic function* :

$$f(x)=\sum_{i=1}^{n/2}(100(x_{2i}-x_{2i-1}^3)^2+(1-x_{2i-1})^2)$$

Starting point $:(-1.2,1,-1.2,1,......)^T$

2.*Non – diagonal function*:

$$(x)=\sum_{i=1}^{n/2}(100(x_i-x_i^3)^2+(1-x_i)^2)$$

Starting point $:(-1,.......................)^T$

3. *Generalized powell function*:

$$f(x)=\sum_{i=1}^{n/4}\begin{matrix}(x_{4i-3}-10x_{4i-2})^2+5(x_{4i-1}-x_{4i})^2+(x_{4i-1}-2x_{4i})^2 \\ +10(x_{4i-9}-x_{4i})^4+(\end{matrix}\; x_{4i-2}-2x_{4i-1}-x_{4i})^2)$$

Starting point: $(3,1,0,1,......................)^T$

4. *Miele function*:

$$f(x)=\sum_{i=1}^{n/4}[\exp(x_{4i-3})-x_{4i-2}]^2+100(x_{4i-2}-x_{4i-1})^6$$

$$+[\tan(x_{4i-1}-x_{4i})]^4+x_{4i-3}^8+(x_{4i}-1)^2$$

Starting point: $(1, 2, 2, 2,................)^T$

5. *Welfe function*:

$$f(x)=(-x_1(3-x_1/2)+2x_2-1)^2+$$

$$\sum_{i=1}^{n-1}(x_{i-1}-x_i(3-x_i(3-x_i/2)+2x_{i+1}-1)^2+(x_{n+1}-x_n(3x_n/2-1)^2$$

Starting point: $(-1, ...)^T$

6. Cantrell *function*:

$$f(x)=\sum_{i=1}^{n/4}[\exp(x_{4i-3})-x_{4i-2}]^4+100(x_{4i-2}-x_{4i-1})^6+$$

$$[\tan^{-1}(x_{4i-1}-x_{4i})]^4+x_{4i-3}^8$$

starting point: $(1, 2, 2, 2,................)^T$

7. *Rosenbrock function*:

$$f(x)=\sum_{i=1}^{n/2}(100(x_{2i}-x_{2i-1}^2)^2+(1-x_{2i-1})^2)$$

Starting point: $(-1.2,1,-1.2,1,......)^T$

8. *Generalized wood function*:

$$f(x)=\sum_{i=1}^{n/4}\begin{matrix}4(x_{4i-2}-x_{4i-3}^2)^2+(1-x_{4i-3})^2+90(x_{4i}-x_{4i-1}^2)^2+(1-x_{4i-1})^2+ \\ 10.1((x_{4i-2}-1)^2+(x_{4i}-1)^2+19.8((x_{4i-2}-1)+(x_{4i}-1))\end{matrix}$$

Starting point: $(-3,-1,-3,-1,.......................)^T$

References

[1] Al-Bayati, A.Y. and Al-Salih, M.S.,(2008):New self-scaling VM-algorithm for non–convex optimization, The 2nd conference Mathematical science ICMS 2008,Allepo University, Syria, pp. 225-237.
[2] AL-Baali M. (1993): Variation Quasi-Newton Methods for unconstrained optimization, J. optimization theory and applications ,Vol.77, pp.127-143.
[3] AL-Baali M. and Fletcher R. (1985): Variation Methods for Non-Linear Least Squares, J. of the Operational Research Society, Vol.36, pp. 405-421.
[4] Greenstadt, J. (1970): Variations on Variable Metric methods Math. Comp. Vol. 24, pp.1-22 .
[5] Oren, S.S. (1974): Self-Scaling Variable Metric Algorithm . Part II, Management Science 20,Vol . (5), pp.862-874.
[6] Oren, S.S. and Luenberger, D. (1974). Self-caling Variable Metric Algorithm. Part I, Management Science 20, Vol. (5), pp.845-862.
[7] Fletcher, R. (1987): Practical Methods of Optimization .John Wiley and Sons, ChiChester (New York) .
[8] Osman G., Filiz G. and Olena S, (2006): Variational Problems In Quasi- Newton Methods Department of Math. and Statistics, University of Maryland Baltimore, County, Maryland 21250, USA, pp.1-22 .

التحديث ألتغييري العام للمتري المتغير في الامثلية غير المقيدة

عباس يونس البياتي[1] ، باسم عباس[1] ، زياد محمد عبدالله[2]

[1]قسم الرياضيات ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل ، الموصل ، العراق

[2]قسم الحاسبات ، كلية علوم الحاسوب والرياضيات ، جامعة تكريت ، تكريت ، العراق

(تاريخ الاستلام: 11 / 5 / 2010 ---- تاريخ القبول: 26 / 10 / 2011)

الملخص

في هذا البحث اقترحنا تغير عام للمتري المتغير لحل مسائل الامثلية غير المقيدة . النتائج العددية اثبت كفاءة الطريقة الجديدة بالاعتماد في المقارنة مع طريقة BFGS القياسية .